MPG analysis, automatic or manual?

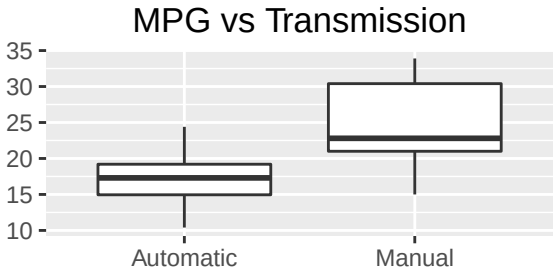
Tom van Dienst

Executive summary

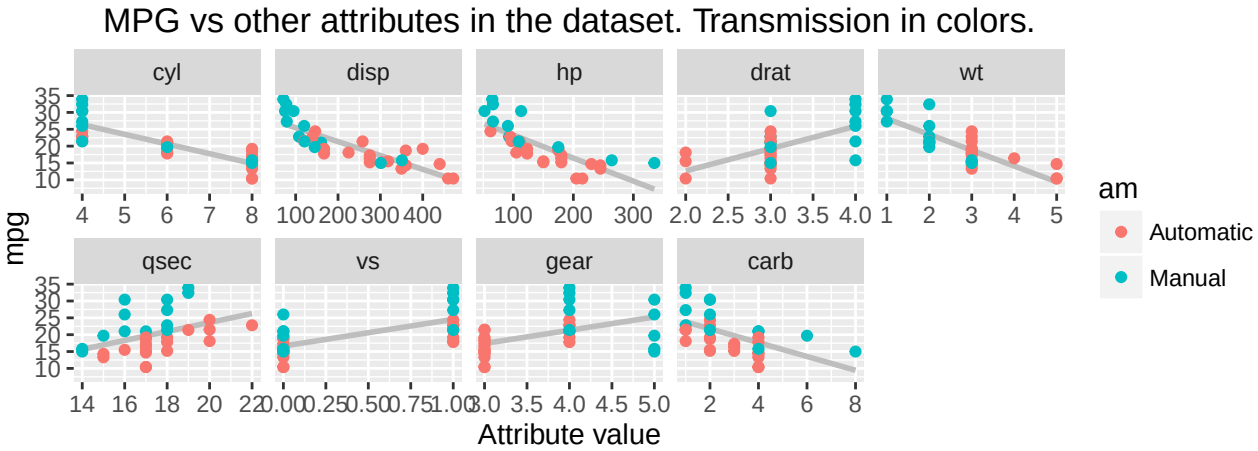
Taking several attributes in account; MPG can be affected by the weight and the transmission of the car. Manual transmission is better for lighter cars and automatic functions better for heavy vehicles.

Main analysis

We make a small alteration to our dataset. The standard variables of am are 0 (automatic) and 1(manual), we'll rename those to their actual descriptions to prevent confusion. Let's take a look at the mpg for automatic vs manual transmission.



It's clear that if you just look at the transmission, an automatic transmission has a much lower MPG than a manual one. But of course there are other attributes which might have something to do with this. Let's take a look at the complete set. We'll see the various measures vs mpg with the transmission as a color. We'll also include a regression line for these variables.



So what do we see? first of all we see that most of the heavier vehicles tend to have an automatic transmission. This seems to match with the displacement and the number of cylinders. Displacement is the total volume of the cylinders, so this is not very surprising. A higher displacement, and weight have a negative influence on the MPG. One value with outliers is the HP, it seems there are some manuals that have a higher horsepower than automatics. This doesn't always match up with the weight. It seems there are some cars with a low weight but high horsepower. One more attribute of interest is 1/4 mile time (qsec), it seems they follow the same slope but have a completely different intercept.

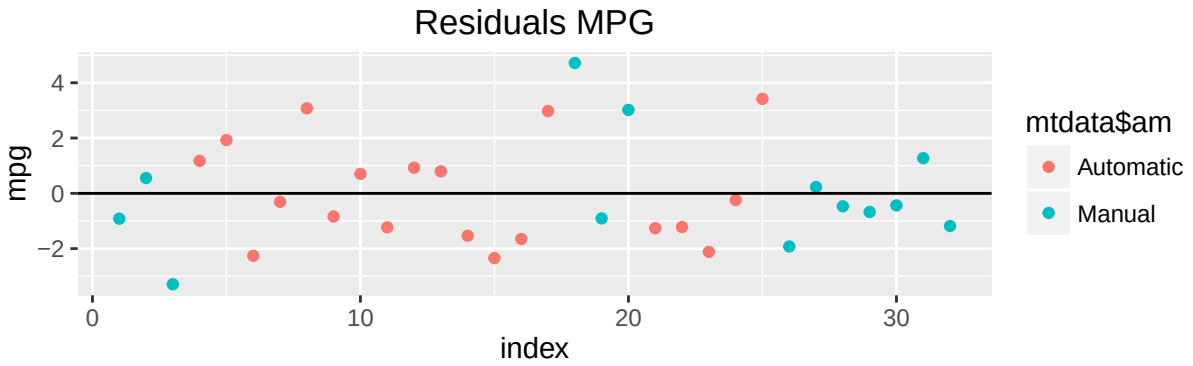
So what will we assume for this model? Our analysis shows that mpg is influenced by weight, displacement, horsepower and qsec. Weight and displacement have a correlation of 0.89 and is considered highly related

so it's enough to just take weight. We'll set up several models and run them through ANOVA to see what influences we can find. We'll start with weight since it has such a big impact on the mpg, next we'll add hp and qsec.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt * am
## Model 2: mpg ~ wt * am + hp * am
## Model 3: mpg ~ wt * am + hp * am + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 188.01
## 2     26 135.90  2    52.108 5.6915 0.009184 **
## 3     25 114.44  1    21.459 4.6877 0.040126 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the analysis shows, a combination of all three is most succesful. Every addition of an attribute decreases our variance significantly.

Let's take a look at the residuals in our model:



The remaining variance seems to be evenly distributed with one outlier in the top. What are our coefficients and confidence intervals?

	Estimate	Pr(>|t|)	2.5 %	97.5 %	Description
(Intercept)	12.225	0.181	-6.093	30.544	Intercept for automatic
wt	-2.512	0.013	-4.437	-0.586	Weight
amManual	14.876	0.001	6.665	23.088	Intercept difference for manual
hp	-0.011	0.553	-0.05	0.027	Horsepower
qsec	0.89	0.04	0.043	1.738	1/4 mile time
wt:amManual	-5.212	0.013	-9.24	-1.183	Weight covariate difference for manual
amManual:hp	0.015	0.441	-0.024	0.054	HP covariate difference for manual

There's definitely a difference between automatic and manual. The model for manual cars shows that they start with a higher initial MPG. But (with all other variables remaining equal) an increase in 1000 lbs for an automatic results in a 2.5 decrease in MPG. The same increase in weight for a manual car would result in $2.5 + 5.2 = 7.7$ decrease in MPG. The other variable of interest is the 1/4 mile time, but this variable does not have a seperate covariate for manual cars and applies for both transmissions. These covariates all have confidence intervals that do not contain 0, and are therefore significant enough to be mentioned. Other variables do not have enough effect.

So what can we conclude? If you're driving a small car, you're beter off with a manual transmission. But

when you're driving a heavy car, you're better off switching to an automatic.

Appendix

Original files

Original files can be found at my github page: **https://github.com/Tothalvadi/Regression-Models**

Setting up the dataset

```
mtdata  <- mutate(mtcars
                  , am = factor(ifelse(am ==  0, "Automatic", "Manual"))
)
```